

ABSTRACT OF THE DISCLOSURE

Several embodiments, among others, provided in the present disclosure teach a filtering of email messages for spam based on phonetic equivalents of words found in the email message. In some embodiments, an email message having a word is received, and
5 a phonetic equivalent of the word is generated. Thereafter, the phonetic equivalent of the word is tokenized to generate a token representative of the phonetic equivalent. The generated token is then used to determine a spam probability.